

Copyright
by
Min Fu
2013

The Report Committee for Min Fu
Certifies that this is the approved version of the following report:

**Dynamic modeling approach to forecast the term
structure of government bond yields**

APPROVED BY
SUPERVISING COMMITTEE:

Supervisor:

Carlos M. Carvalho

James G. Scott

**Dynamic modeling approach to forecast the term
structure of government bond yields**

by

Min Fu, B.S.

REPORT

Presented to the Faculty of the Graduate School of
The University of Texas at Austin
in Partial Fulfillment
of the Requirements
for the Degree of

Master of Science in Statistics

THE UNIVERSITY OF TEXAS AT AUSTIN

May 2013

Acknowledgments

I would like to thank Dr. Carlos Carvalho for his dedicated supervision from hunting for the topic to the writing of this report. I also want to thank Dr. James Scott for his advice and his course which provides me a solid theoretical background for Bayesian statistics. Additional thank goes to my husband, who supports me all the time.

Dynamic modeling approach to forecast the term structure of government bond yields

Min Fu, M.S.Stat

The University of Texas at Austin, 2013

Supervisor: Carlos M. Carvalho

Since arbitrage-free is a desirable theoretical feature in a healthy financial market, many efforts have been made to construct arbitrage-free models for yield curves. However, little attention is paid to review if such restriction will improve yield forecast. We evaluate the importance of arbitrage-free restriction on dynamic Nelson-Siegel term structure when forecasting yield curves. We find that it doesn't help. We also compare these two Nelson-Siegel dynamic models with a benchmark dynamic model and show that Nelson-Siegel structure improve forecasts for long-maturity yields.

Table of Contents

Acknowledgments	iv
Abstract	v
List of Tables	vii
List of Figures	viii
Chapter 1. Introduction	1
Chapter 2. Characteristics of yield curves	3
Chapter 3. Literature Review	6
Chapter 4. Dynamic Nelson-Siegel Model	8
4.1 Nelson-Siegel term structure	8
4.2 Dynamic Nelson-Siegel Model	10
4.3 Estimating latent states	11
4.4 Estimating parameters	12
4.5 Forecasting	13
Chapter 5. Arbitrage-free restriction	15
Chapter 6. Results	17
6.1 Data	17
6.2 Model estimation	17
6.3 Out-of-sample forecasting	22
Chapter 7. Discussion	26
Bibliography	28

List of Tables

6.1	Parameter estimates of DNS model	19
6.2	Parameter estimates of AFDNS model	19
6.3	Summary statistics of in-sample fit	23
6.4	Out-of-sample forecast RMSE	25

List of Figures

2.1	Yield curves from Jan 1985 to Dec 2000 constructed by maturities of 3, 6, 9, 12, 15, 18, 21, 24, 30, 36, 48, 60, 72, 84, 96, 108 and 120 months	4
2.2	Mean yield curve across time at maturities of 3, 6, 9, 12, 15, 18, 21, 24, 30, 36, 48, 60, 72, 84, 96, 108 and 120 months. Sample period is from Jan 1985 to Dec 2000.	5
4.1	Three factor loadings in Nelson-Siegel model, which are $1, \frac{1-e^{-\lambda\tau}}{\lambda\tau}, \frac{1-e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau}$ respectively. Set $\lambda = 0.0609$	9
6.1	diagnosis plot and posterior distribution of λ	18
6.2	Comparison of estimated level, slope and curvature across models	20
6.3	The yield-adjustment term for different maturities	22

Chapter 1

Introduction

Interest rates or yield curve forecasting is important for bond portfolio management, derivatives pricing, risk management, monetary policy making, capital goods purchase and fiscal debt structure. Yield curve shows the relationship between bond yield and different maturities or terms, therefore it is also known as term structure. Researchers have proposed a large variety of models to investigate dynamics of yield curve and produce forecast. Dynamic Nelson-Siegel model proposed by Diebold and Li(2006) proves to be a success. However, it doesn't ensure arbitrage-free, which is an important theoretical assumption that rules out opportunities for risk-free arbitrage across maturities in well-organized market. Therefore, Christensen et al.(2007) proposed an affine arbitrage-free class of Nelson-Siegel term structure models.

Our purpose of this paper is twofold: 1) review the validity of dynamic Nelson-Siegel model and arbitrage-free dynamic Nelson-Siegel model from theoretical perspective; 2) using real data to investigate the role of arbitrage-free restriction to the dynamic Nelson-Siegel model.

In the following chapters, we will first discuss the characterization of yield curves in Chapter 2 and review the different modeling approaches hav-

ing been proposed in Chapter 3. Then we will describe in details the dynamic modeling approach and how it applies to Nelson-Siegel term structure. Chapter 6 will give empirical results regarding two models, both in-sample fit and out-of-sample forecast, and evaluate the role of arbitrage-free. Finally we conclude in Chapter 7.

Chapter 2

Characteristics of yield curves

We do not observe yields directly. A popular approach to construct yield from observed bond prices is called "unsmoothed Fama-Bliss"(1987) method. Figure 2.1 shows the unsmoothed Fama-Bliss yield curves across time. We see the shapes of these yield curves differ greatly, some are upward sloping and some are downward. Also the increasing or decreasing rate of the slopes differ a lot.

In general terms, we expect that yields increase with maturity because of the intuition that lenders demand higher interest rates for longer-term loans as compensation for the greater risk associated with them, in comparison to short-term loans. We call it an "inverted yield curve" if long-term yields fall below short-term yields, which is generally regarded as a harbinger of recession. Figure 2.2 shows that the mean yield curve has an upward slope with decreasing rate.

Even though yield curve may have different shapes, earlier observation recognized yields can be well explained by three components. See Litterman and Scheinkman(1991), Bliss(1997) and Duffee(2002). Joslin et al.(2010) state that:

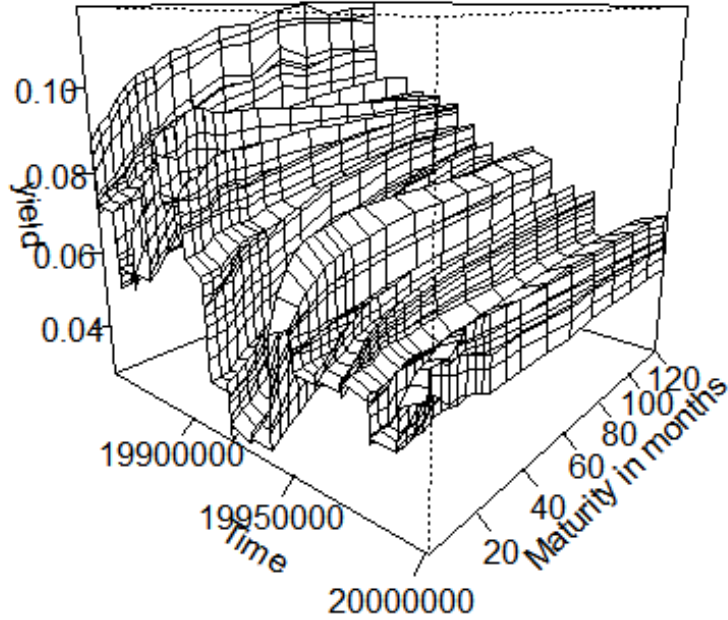


Figure 2.1: Yield curves from Jan 1985 to Dec 2000 constructed by maturities of 3, 6, 9, 12, 15, 18, 21, 24, 30, 36, 48, 60, 72, 84, 96, 108 and 120 months

The cross correlations of bond yields are well described by a low-dimensional factor model in the sense that the first three principal components of bond yields ... explain well over 95 percent of their variation ... Very similar three-factor representations emerge from arbitrage-free dynamic term structure models.

These three components correspond to level, slope and curvature (the 10-year yield, the 10Y-6M spread and a 6M+10Y-2*2Y butterfly spread empirically).

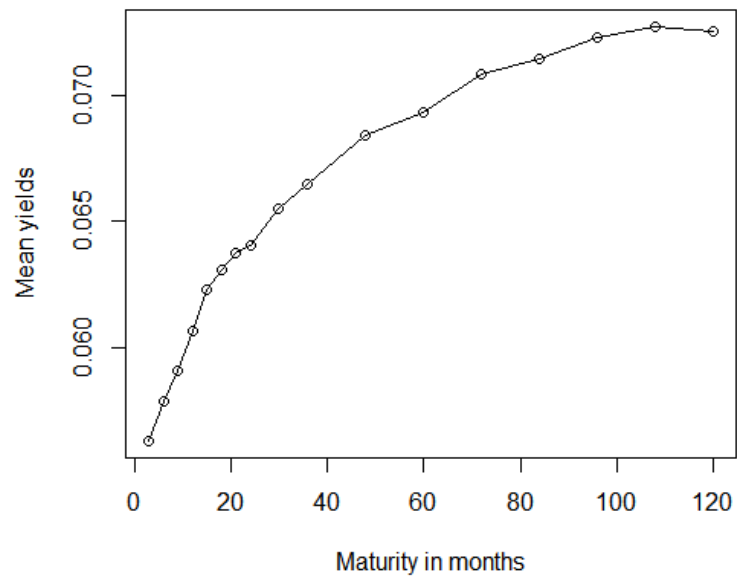


Figure 2.2: Mean yield curve across time at maturities of 3, 6, 9, 12, 15, 18, 21, 24, 30, 36, 48, 60, 72, 84, 96, 108 and 120 months. Sample period is from Jan 1985 to Dec 2000.

Chapter 3

Literature Review

Over the last 30 years, researches have successfully proposed many theoretical models to investigate the term structure. A vast financial literature contributes to arbitrage-free model. Such theoretical restriction rules out opportunities for no-risk arbitrage at any point in time, which is important for pricing derivatives. Prominent contributions starts from Vasicek(1977), Cox et al.(1985). Duffie and Kan(1996) proposed the affine version of arbitrage-free model, where yields are linear function of latent factors and factor loadings can be calculated from a system of differential equations.

Recently, researchers become more interested in the dynamics of yield curve and try to produce good forecast. Duffee(2002) shows that the affine arbitrage-free model may produce poor forecast, no better than forecasting by simple random walk model. Besides, the estimation of such kind of model is hard due to the generality of the model and too many parameters need to be estimated. Diebold and Li(2006) propose a dynamic model of term structure based on Nelson and Siegel(1987) exponential components framework, which represents yield curve by three factors and imposes structure on factor loadings. Unlike the affine arbitrage-free model, in which we need to estimate both

the latent factors and factor loadings, here we just need to fit a linear regression at any point in time. Diebold and Li introduce dynamics to three factors to capture the evolution of yield curve and their forecasts beat other popular models. However, the Nelson-Siegel framework has less strong theoretical foundation even though it is popular among financial market practitioners. Christensen et al.(2007) show that the model can be interpreted as a specific type of the affine arbitrage-free model with simple modification.

In the mean time, Joslin et al.(2011) and Duffee(2011) argue that arbitrage-free restriction doesn't help to the term structure estimation and forecasting since three-factor mapping is equivalent to represent cross-sectional yield curve, and arbitrage-free restriction on the factors improves little. Therefore, Duffee(2011) proposes a three-factor dynamic term structure model with no arbitrage-free condition. Three factors are the first three principal components of yield curves, and dynamics are imposed on these principal components.

We will focus on dynamic models based on Nelson-Siegel framework, and use dynamic model with three principal components as comparison.

Chapter 4

Dynamic Nelson-Siegel Model

In this chapter, we will introduce the classic Nelson-Siegel term structure first, then incorporate time dynamics into it and discuss how to estimate this model and construct forecasts.

4.1 Nelson-Siegel term structure

Let's consider the cross-sectional yields first. At any time, we see a set of (Fama-Bliss unsmoothed) yields. Nelson-Siegel(1987) proposed a parsimonious three-component function to model the yield curve, which is

$$y(\tau) = \beta_1 + \beta_2 \left(\frac{1 - e^{-\lambda\tau}}{\lambda\tau} \right) + \beta_3 \left(\frac{1 - e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau} \right), \quad (4.1)$$

where $y(\tau)$ is zero-coupon yield, τ is bond maturity in months.

$\beta_1, \beta_2, \beta_3$ are latent factors. Loading of β_1 is constant 1; loading of β_2 is a decreasing function starting from 1 to 0; loading of β_3 starts from 0, increases and then decreases to 0. The parameter λ controls the decay rate of β_2 and when β_2 achieves maximum.¹Figure 4.1 gives the plot of three loadings.

¹Here we set $\lambda = 0.0609$ which maximizes the loading of β_2 at 30 months. Choice of λ will be discussed in details later.

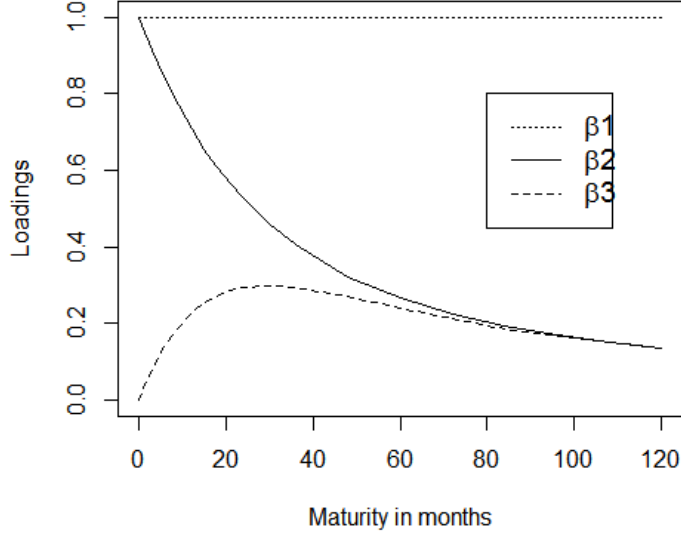


Figure 4.1: Three factor loadings in Nelson-Siegel model, which are 1 , $\frac{1-e^{-\lambda\tau}}{\lambda\tau}$, $\frac{1-e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau}$ respectively. Set $\lambda = 0.0609$.

It is important to notice that three factors can be interpreted as level, slope and curvature. It is easy to show that $y(120) \approx y(\infty) = \beta_1$. An increase in β_1 will inflate yields at different maturities equally, therefore changing the level. Also, $y(120) - y(6) \approx y(\infty) - y(0) = -\beta_2$. An increase in β_2 increases short-term yields by much larger amounts than long-term yields, therefore changing the slope(yield curve becomes steeper). Finally, $y(6) + y(120) - 2 * y(24) = -0.07\beta_2 - 0.31\beta_3$. An increase in β_3 increases medium-term yields much greater than short-term or long-term yields, therefore changing the curvature(yield curve becomes more 'hump-shaped').

4.2 Dynamic Nelson-Siegel Model

Following Diebold and Li(2006), Nelson-Siegel term structure can be interpreted as a dynamic model where $\{\beta_1, \beta_2, \beta_3\}$ are time-varying latent factors corresponding to level, slope and curvature at each time t . The evolution of three β 's are univariate AR(1) processes.²Therefore, dynamic Nelson-Siegel model(DNS) is given by:

$$y_t(\tau) = \beta_{1t} + \beta_{2t}\left(\frac{1 - e^{-\lambda\tau}}{\lambda\tau}\right) + \beta_{3t}\left(\frac{1 - e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau}\right) + \sigma_t,$$

$$\sigma_t \sim N(0, V) \quad (4.2)$$

$$\beta_{it} = a_i + b_i\beta_{i,t-1} + \omega_{it}, \omega_{it} \sim N(0, W_i), i = 1, 2, 3 \quad (4.3)$$

We can rewrite it as

$$y_t(\tau) = F(\tau)\beta_t + \sigma_t, \sigma_t \sim N(0, V), \quad (4.4)$$

$$\beta_t = a + b\beta_{t-1} + \omega_t, \omega_t \sim N(0, W), \quad (4.5)$$

where $F(\tau) = (1, \frac{1-e^{-\lambda\tau}}{\lambda\tau}, \frac{1-e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau})$, $\beta_t = (\beta_{1t}, \beta_{2t}, \beta_{3t})^T$, $a = (a_1, a_2, a_3)^T$, $b = \text{diag}(b_1, b_2, b_3)$, $W = \text{diag}(w_1, w_2, w_3)$ and $V = \text{diag}(v_1, v_2, v_3)$.

We can use Gibbs sampler to estimate parameters and latent states. Given parameters $\{a, b, V, W\}$, we can use Forward Filtering Backwards Sampling(FFBS) to obtain posterior sample of $\{\beta_{1:T}\}$ given data. And given states, we can update parameters easily based on conjugate prior. We sample parameter and states iteratively until achieving convergence.

²Diebold and Li(2006) shows that univariate AR(1) performs better than multivariate VAR(1) process. Also level, slope and curvature are the first three principal components of yield covariance matrix, which gives us some sense that they should be orthogonal.

4.3 Estimating latent states

Assume $\{a, b, V, W\}$ are known, let posterior distribution of β at $t - 1$ be

$$\beta_{t-1}|y_{1:t-1} \sim N(m_{t-1}, C_{t-1}). \quad (4.6)$$

The one-step-ahead predictive distribution of β_t given $y_{1:t-1}$ is

$$\beta_t|y_{1:t-1} \sim N(a_t, R_t), \quad (4.7)$$

where $a_t = a + bm_{t-1}$, $R_t = bC_{t-1}b^T + W$.

The one-step-ahead predictive distribution of y_t given $y_{1:t-1}$ is

$$y_t|y_{1:t-1} \sim N(f_t, Q_t), \quad (4.8)$$

where $f_t = Fa_t$, $Q_t = FR_tF^T + V$.

The posterior distribution β at time t is

$$\beta_t|y_{1:t} \sim N(m_t, C_t), \quad (4.9)$$

where $m_t = a_t + R_tF_t^TQ_t^{-1}e_t$, $e_t = y_t - f_t$; $C_t = R_t - R_tF_t^TQ_t^{-1}F_tR_t$.

Therefore given a starting point $\beta_0|y_0 \sim N(m_0, C_0)$, we can filter forward and get $P(\beta_t|y_{1:t})$ at each time period. To draw sample from $P(\beta_{1:T}|y_{1:T})$, we notice that by conditional independence assumption we have

$$P(\beta_{1:T}|y_{1:T}) = P(\beta_T|y_{1:T})P(\beta_{T-1}|\beta_T, y_{1:T}) \dots P(\beta_1|\beta_2, \dots, \beta_T, y_{1:T}) \quad (4.10)$$

$$= P(\beta_T|y_{1:T}) \prod_{t=T-1}^{t=1} P(\beta_t|\beta_{t+1}, y_{1:t}). \quad (4.11)$$

We already know $P(\beta_T|y_{1:T})$ during filtering. And it can be shown that

$$\beta_t|\beta_{t+1}, y_{1:t} \sim N(h_t, H_t), \quad (4.12)$$

where $h_t = m_t + C_t G_{t+1}^T R_{t+1}^{-1} (\beta_{t+1} - a_{t+1})$, $H_t = C_t - C_t G_{t+1}^T R_{t+1}^{-1} G_{t+1} C_t$.

Therefore by eqn (4.11) we can draw β_T first, then draw β_t backwards from $N(h_t, H_t)$, $t = T - 1, \dots, 1$ to get a joint sample from $P(\beta_{1:T}|y_{1:T})$.

4.4 Estimating parameters

Assume $\{\beta_{1:T}\}$ are known. As the prior, we assume the AR parameters are i.i.d. Gaussian and variance terms are Inverse-Gamma

$$a_i \sim N(\alpha_0, \tau_0), i = 1, 2, 3 \quad (4.13)$$

$$b_i \sim N(\phi_0, \sigma_0), i = 1, 2, 3 \quad (4.14)$$

$$w_i \sim IG(c/2, d/2), i = 1, 2, 3 \quad (4.15)$$

$$v_j \sim IG(e/2, f/2), j = 1, \dots, m \quad (4.16)$$

The full conditional distribution of w_i is

$$w_i | \dots \sim IG\left(\frac{c+T}{2}, \frac{d+SS_{\beta,i}}{2}\right), i = 1, 2, 3 \quad (4.17)$$

where $SS_{\beta,i} = \sum_{t=1}^T (\beta_{i,t} - (a_i - b_i \beta_{i,t-1}))^2$.

The full conditional distribution of v_j is

$$v_j | \dots \sim IG\left(\frac{e+T}{2}, \frac{f+SS_{y,j}}{2}\right), j = 1, \dots, m \quad (4.18)$$

where $SS_{y,j} = \sum_{t=1}^T (y_{t,j} - (F\beta_t)_j)^2$.

The full conditional distribution of a_i is

$$a_i | \dots \sim N(\alpha_{1,i}, \tau_{1,i}), i = 1, 2, 3 \quad (4.19)$$

where $\tau_{1,i} = \left[\frac{1}{\tau_0} + \frac{T}{w_i} \right]^{-1}$, $\alpha_{1,i} = \tau_{1,i} \left[\frac{\alpha_0}{\tau_0} + \frac{\sum_{t=1}^T (\beta_{i,t} - b_i \beta_{i,t-1})}{w_i} \right]$.

The full conditional distribution of b_i is

$$b_i | \dots \sim N(\phi_{1,i}, \sigma_{1,i}), i = 1, 2, 3 \quad (4.20)$$

where $\sigma_{1,i} = \left[\frac{1}{\sigma_0} + \frac{\sum_{t=1}^T \beta_{i,t-1}^2}{w_i} \right]^{-1}$, $\phi_{1,i} = \sigma_{1,i} \left[\frac{\phi_0}{\sigma_0} + \frac{\sum_{t=1}^T (\beta_{i,t} - a_i) \beta_{i,t-1}}{w_i} \right]$.

Therefore, we can easily sample from these posterior distributions to get estimates of these unknown parameters.

4.5 Forecasting

With data $y_{1:T}$, we are interested in forecasting future yield values. For dynamic model, it is easy and straightforward to compute forecast. One-step-ahead forecast can be naturally computed during forward filtering. For k-step-ahead forecast, we just propagate states from T to $T + k$ by evolution equation and then get predictive distribution of y_{T+k} by observation equation.

It can be shown that predictive distribution of y_{T+k} given $y_{1:T}$ is

$$y_{T+k} | y_{1:T} \sim N(f_{T+k}, Q_{T+k}), \quad (4.21)$$

where $f_{T+k} = F a_{T+k}$, $Q_{T+k} = F R_{T+k} F^T + V$. Here a_{T+k} , R_{T+k} are the mean and variance of predictive distribution of β_{T+k} given $y_{1:T}$, and can be computed

recursively by

$$a_{T+k} = a + ba_{T+k-1}, \quad (4.22)$$

$$R_{T+k} = bR_{T+k-1}b^T + W. \quad (4.23)$$

Chapter 5

Arbitrage-free restriction

Arbitrage-free restriction ensures consistency between the dynamic evolution of yields over time and the shape of the yield curve at a given time point. It is a desirable property in quantitative finance which implies an equivalent-martingale measure.

Dynamic Nelson-Siegel model introduced above doesn't ensure arbitrage-free. But Christensen et al.(2007) showed that arbitrage-free restriction can be imposed by just adding a yield-adjustment term to the original model. Arbitrage-free dynamic Nelson-Siegel model(AFDNS) is given by

$$y_t(\tau) = C(\tau) + F(\tau)\beta_t + \sigma_t, \sigma_t \sim N(0, V), \quad (5.1)$$

$$\beta_t = a + b\beta_{t-1} + \omega_t, \omega_t \sim N(0, W), \quad (5.2)$$

here C is an yield-adjustment term.

Since arbitrage-free imposes additional restrictions on cross-sectional mapping (see Duffie and Kan(1996)), AR parameters in the evolution equation are no longer free parameters. To be more specific, assuming yields are semi-martingale processes under pricing or Q -measure, express the restrictions

under real-world P-measure

$$b = \exp(-K^P \Delta t), \quad (5.3)$$

$$W = \int_0^{\Delta t} \exp(-K^P s) \Sigma \Sigma^T \exp(-(K^P)^T s) ds, \quad (5.4)$$

where Δt is the interval between two observations of states, K^P is the mean-reversion matrix under P-measure, Σ is the volatility matrix of latent states.

As in Chapter 4, we assume three factors are independent, then the mean-reversion matrix K^P and the volatility matrix Σ are diagonal. We write Σ as $\text{diag}(\sigma_1, \sigma_2, \sigma_3)$. The yield-adjustment term can be computed as

$$\begin{aligned} C(\tau) = & \sigma_1^2 \frac{\tau^2}{6} + \sigma_2^2 \left[\frac{1}{2\lambda^2} - \frac{1 - e^{-\lambda\tau}}{\lambda^3\tau} + \frac{1 - e^{-2\lambda\tau}}{4\lambda^3\tau} \right] + \sigma_3^2 \left[\frac{1}{2\lambda^2} + \frac{1}{\lambda^2} e^{-\lambda\tau} - \right. \\ & \left. - \frac{1}{4\lambda} \tau e^{-2\lambda\tau} - \frac{3}{4\lambda^2} e^{-2\lambda\tau} - \frac{2(1 - e^{-\lambda\tau})}{\lambda^3\tau} + \frac{5(1 - e^{-2\lambda\tau})}{8\lambda^3\tau} \right]. \end{aligned} \quad (5.5)$$

We can still use Gibbs Sampler to estimate parameters and latent states. A few things to notice are that we need to ensure b is positive and we need to calculate Σ based on estimated W so that we can get the yield-adjustment term. For the first one, we can still assign normal prior to b , just use different hyperparameters to ensure the samples will be positive. For the second one, since Σ is diagonal under assumption, the integral can actually be written as simple function form and the inverse can be easily derived.

Chapter 6

Results

6.1 Data

We use end-of-month, unsmoothed Fama-Bliss(1987) zero-coupon yields at maturities of 3,6,9,12,15,18,21,24,30,36,48,60,72,84,96,108, and 120 months. The data is from 1985.01 to 2000.12. It is the same data set used by Diebold and Li(2006).

6.2 Model estimation

We use yield data in original basis, and pick the hyperparameters of parameter priors as $\alpha_0 = 0, \tau_0 = 5e - 8, c/2 = 3, d/2 = 1e - 4, e/2 = 3, f/2 = 1e - 6$, so that noise-signal ratio is 0.01. We start Kalman filter with $m_0 = 0, C_0 = 1$. Note that we do not assume the stationarity of the states.

First, we investigate if it is appropriate to set λ as 0.0609, which maximizes the medium-term factor at 30 months. We check it on DNS model using Gibbs Sampler. We use conjugate priors for model parameters, independence Metropolis for λ and Kalman filter for latent states. Figure 1 shows the Gibbs sample of λ . The acceptance rate is 58.11% and posterior mean is 0.0609. Therefore, in the following analysis, we will use λ as 0.0609 since it will sim-

plify the non-linear Nelson-Siegel term structure to a linear function of latent factors. So we can use dynamic model approach discussed in previous chapters to do estimation and forecasting.

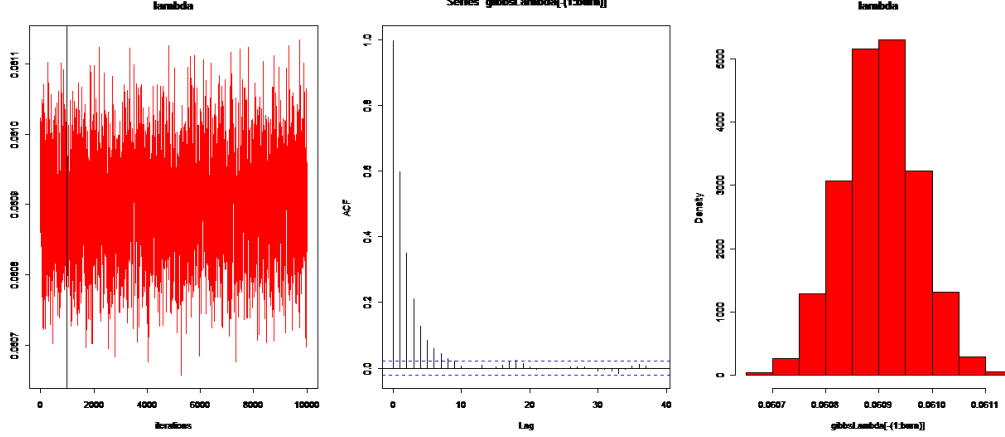


Figure 6.1: diagnosis plot and posterior distribution of λ

Table 6.1 and 6.2 give MCMC posterior estimates of AR parameters and variance for DNS and AFDNS models respectively, along with their Monte Carlos standard errors. Diagnostic plots (not shown) indicate convergence of MCMC chain. We burn in the first 1000 sample out of 10000 total.

Figure 6.2 plots estimated $\beta_{1t}, \beta_{2t}, \beta_{3t}$ for DNS model, the empirical level, slope, curvature as well as the first three principal components of covariance matrix of yield data. We didn't plot estimated factors for AFDNS because they are very similar to DNS estimates. The plot show that factors in Nelson-Siegel model have empirical meaning as level, slope and curvature. The correlation between the estimated $\beta_{1t}, \beta_{2t}, \beta_{3t}$ and empirical level, slope and curvature are $\text{cor}(\hat{\beta}_{1t}, L_t)=0.970$, $\text{cor}(\hat{\beta}_{2t}, S_t)=-0.996$, $\text{cor}(\hat{\beta}_{3t}, C_t)=-0.966$,

Table 6.1: Parameter estimates of DNS model

	a	b	\sqrt{W}
1	2.15e-04 (2.04e-06)	9.90e-01 (2.98e-05)	3.13e-03 (1.96e-06)
2	-1.99e-06 (2.02e-06)	9.89e-01 (9.57e-05)	3.21e-03 (2.01e-06)
3	-7.87e-05 (2.14e-06)	9.15e-01 (3.22e-04)	6.47e-03 (5.42e-06)

Note: We present estimates of AR parameters and standard error of states. Monte Carlo standard errors of parameters are given in parentheses. λ is set to be 0.0609 (maximizes the loading of medium-term factor at 30 months). Matrix of b and W are diagonal, so we just give estimates of their diagonal elements.

Table 6.2: Parameter estimates of AFDNS model

	a	b	\sqrt{W}
1	2.13e-04 (2.04e-06)	9.90e-01 (2.99e-05)	3.14e-03 (1.96e-06)
2	-6.14e-07 (2.02e-06)	9.89e-01 (9.55e-05)	3.22e-03 (1.98e-06)
3	-7.96e-05 (2.15e-06)	9.11e-01 (3.30e-04)	6.58e-03 (5.29e-06)

Note: We present estimates of AR parameters and standard error of states. Monte Carlo standard errors of parameters are given in parentheses. λ is set to be 0.72 (maximizes the loading of medium-term factor at 2.5 years). Note that we transform the yield data to year basis. Matrix of b and W are diagonal, so we just give estimates of their diagonal elements.

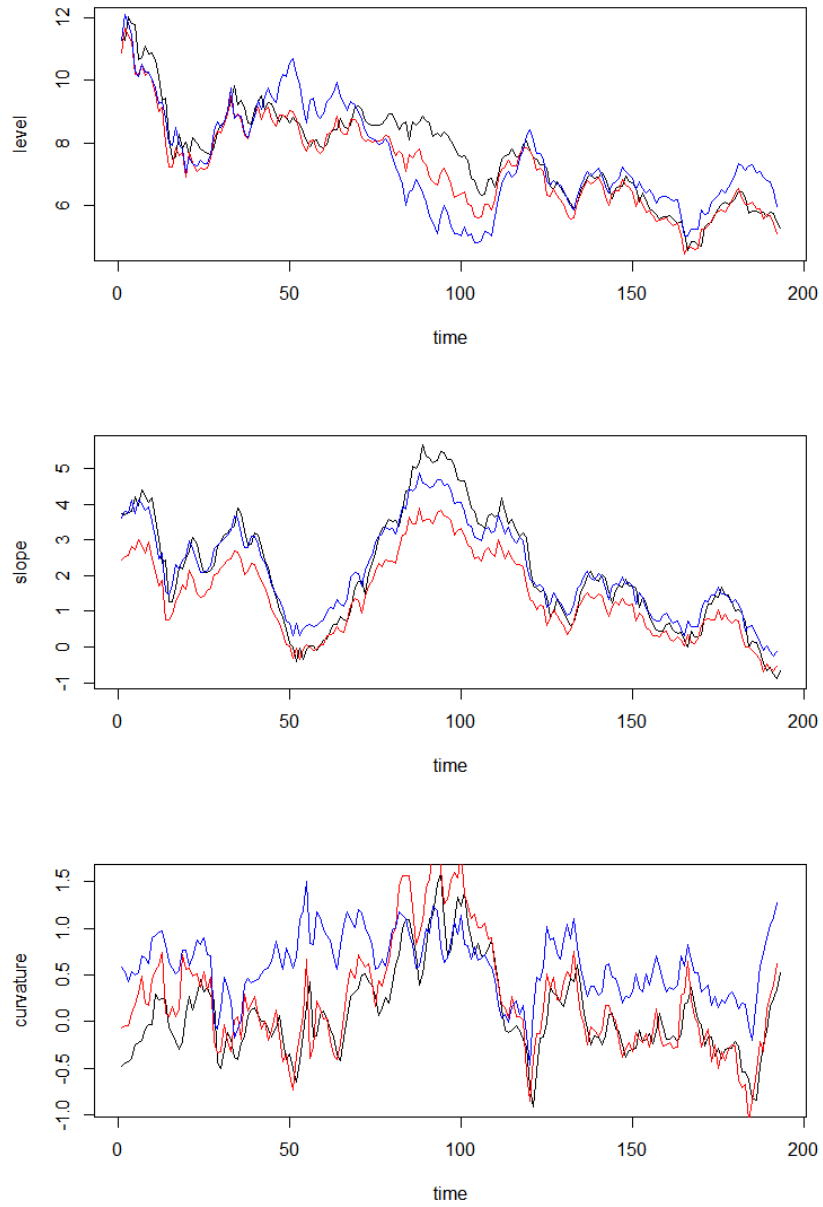


Figure 6.2: Comparison of estimated level, slope and curvature across models
 Note: black line: $\hat{\beta}_{1t}$, $-\hat{\beta}_{2t}$, $-0.3 * \hat{\beta}_{3t}$ respectively; red line: empirical level, slope and curvature; blue line: three principle components

where (I_t, S_t, C_t) denote empirical level, slope and curvature. The correlation between three principal components and empirical level, slope and curvature are $\text{cor}(\text{1st pc}, L_t) = -0.901$, $\text{cor}(\text{2nd pc}, S_t) = 0.994$, $\text{cor}(\text{3rd pc}, C_t) = 0.637$.

From model estimation we also see $\hat{\beta}_{1t}$ and $\hat{\beta}_{2t}$ are more persistent comparing to $\hat{\beta}_{3t}$. Variance of three states are of the same scale, while for the third factor it is a little higher, indicating it is the hardest to predict. Duffee(2011) used dynamic model based on principal components and reached similar conclusion.

For AFDNS model, we convert matrix b to get K^P

$$K^P = -12\log(b) = \begin{pmatrix} 0.1180 & 0 & 0 \\ 0 & 0.1268 & 0 \\ 0 & 0 & 1.1213 \end{pmatrix}. \quad (6.1)$$

We also convert variance matrix W into volatility matrix Σ by $\sigma_i^2 = \frac{24\log(b_i)w_i}{b_i^2 - 1}, i = 1, 2, 3$,

$$\Sigma = \begin{pmatrix} 0.0095 & 0 & 0 \\ 0 & 0.01069 & 0 \\ 0 & 0 & 0.02395 \end{pmatrix}. \quad (6.2)$$

Figure 6.3 plots the yield-adjustment term calculated based on Σ . All the adjustments are negative and get larger for longer maturities. Therefore this term has greater influence on long-term yields.

Table 6.3 compares the in-sample fit for two models. We see that two perform similar before 60 months, but DNS shows better in-sample fit for long-term yields which can be explained by the increased adjustment term.

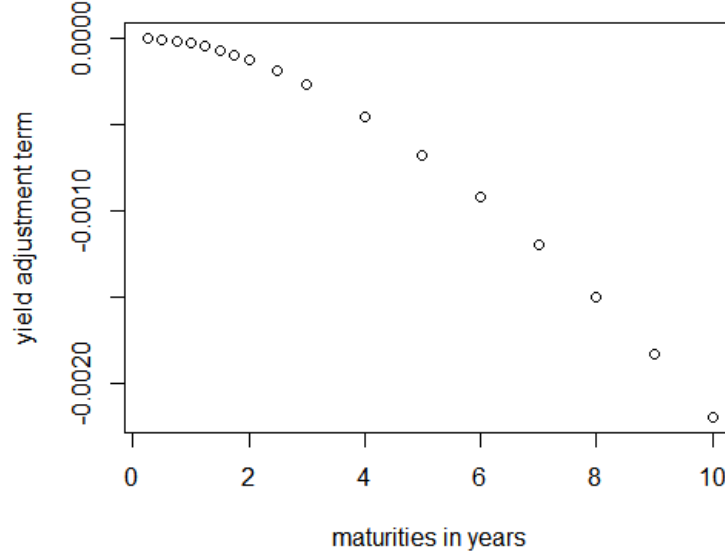


Figure 6.3: The yield-adjustment term for different maturities

From another perspective, since additional restrictions are imposed on AFDNS model, it can not fit cross-sectional yields as free as DNS model. However, we are more interested in predicting future trends. If poor in-sample fit can be offset by better forecast, it is worthwhile. So we go on to evaluate out-of-sample forecast performance.

6.3 Out-of-sample forecasting

In this section, we will investigate out-of-sample forecast accuracy of DNS and AFDNS models. We will use forecasting based on PCA as benchmark.

Table 6.3: Summary statistics of in-sample fit

Maturity in months	DNS		AFDNS	
	mean	RMSE	mean	RMSE
3	-3.05	13.65	-3.15	14.15
6	-1.74	4.92	-1.69	5.15
9	-2.37	5.17	-2.16	4.93
12	2.00	8.20	2.40	8.37
15	7.30	9.14	7.91	9.74
18	5.99	7.33	6.83	6.37
21	3.92	5.30	5.01	6.37
24	-1.36	4.81	0.006	4.94
30	-0.37	1.96	1.62	2.12
36	-2.55	4.24	0.15	3.37
48	-1.14	5.98	3.21	6.64
60	-5.13	7.88	1.10	6.10
72	0.62	8.52	8.95	12.37
84	-0.65	6.24	10.00	11.80
96	2.06	4.18	15.28	15.68
108	1.92	4.17	17.93	18.28
120	-3.32	7.84	15.76	17.27

Note: We present the mean and RMSE for 17 different maturities. All numbers are measured in basis points.

We construct k-month-ahead forecasts for yields with maturities of 3,12,36,60 and 120 months, and forecast horizons of $k = 1, 6$ and 12 months. We estimate and forecast recursively, using data from 1985.01 to the time when forecast is made, then adding one month of data and conducting another forecast. For one-month-ahead forecast, the forecasts are made from 1994.12 to 2000.11. There are 72 forecasts in all. For 6-month-ahead forecast, the forecasts are made from 1994.12 to 2000.06, 67 forecasts in all. For

12-month-ahead forecast, the forecasts are made from 1994.12 to 1999.12, 61 forecasts in all.

PCA model is to first conduct principal components analysis to covariance matrix of yield and extract the largest three eigenvalues and corresponding eigenvectors. In model estimation part, we have showed that these three principal components (by projecting yield data to three-eigenvector space) can be interpreted as level, slope and curvature. Then we fit an univariate AR(1) model to produce forecasts of them respectively and project them back to original space to generate yield forecast.

Table 6.4 presents the out-of-sample performance for PCA, DNS and AFDNS model. First, we see the results for DNS and AFDNS are very close which means the yield adjustment term has little effect on forecasting. In other words, arbitrage-free restriction doesn't improve forecast very much for Nelson-Siegel term structure. We also see PCA performs better for forecasting short-maturity yields but bad at long-maturity yields because of no shape constrain of discount curve. But Nelson-Siegel model ensures that discount curve starts from one at zero maturity and approaches zero at infinite maturity as well as positive forward rates all the time. Therefore, for forecasting yield with large maturity, Nelson-Siegel term structure prove to be successful.

Table 6.4: Out-of-sample forecast RMSE

Model	Forecast		
	1 month	6 months	12 months
	3-month yield		
PCA	15.78	37.63	56.41
DNS	15.81	46.18	80.24
AFDNS	16.71	46.19	80.29
	12-month yield		
PCA	19.73	52.16	67.79
DNS	19.39	58.10	89.35
AFDNS	19.36	57.97	89.44
	36-month yield		
PCA	26.13	68.73	81.66
DNS	24.94	67.50	91.27
AFDNS	24.91	67.31	91.29
	60-month yield		
PCA	29.88	79.68	97.76
DNS	26.47	69.65	92.35
AFDNS	26.47	69.49	92.34
	120-month yield		
PCA	25.02	71.20	96.97
DNS	23.97	61.49	84.52
AFDNS	24.01	61.40	84.59

Note: For each maturity and horizon, we present RMSE of three models. All numbers are measured in basis points.

Chapter 7

Discussion

We have reviewed the dynamic modeling approach built upon Nelson-Siegel term structure and arbitrage-free restriction. We have evaluated the in-sample fit as well as out-of-sample forecasting performance for both models and compare them with principal component dynamic regression proposed by Duffee(2011).

It is shown that arbitrage-free restriction adds little to Nelson-Siegel dynamic model since their forecasting performance is very close. This may be because Nelson-Siegel dynamic model already fulfill arbitrage-free restriction to some extent even if it doesn't explicitly require that, or simply because arbitrage-free restriction doesn't help with forecast. In recent years, in spite of many papers devoted to arbitrage-free model, researchers begin to re-inspect the importance of arbitrage-free restriction and whether it is essential when modeling the yield curve, see Joslin et al.(2011) and Duffee(2011).

We also see Nelson-Siegel term structure predicts better for long-maturity yields comparing to principal component dynamic regression, which implies imposed structure on factor loading is useful to model the trend of yield curves. Another thing to notice is that Nelson-Siegel dynamic model provides forecast

of the smoothed yield curves, while principal component dynamic regression only forecasts yields of given maturities by the data. Therefore, Nelson-Siegel model is preferred due to high flexibility provided.

Bibliography

- [1] Robert R. Bliss. Movements in the term structure of interest rates. *Economic Review*, (Q 4):16–33, 1997.
- [2] Jens H.E. Christensen, Francis X. Diebold, and Glenn D. Rudebusch. The affine arbitrage-free class of nelson-siegel term structure models. *Journal of Econometrics*, 164(1):4–20, September 2011.
- [3] John C Cox, Jr Ingersoll, Jonathan E, and Stephen A Ross. A theory of the term structure of interest rates. *Econometrica*, 53(2):385–407, March 1985.
- [4] Francis X. Diebold and Canlin Li. Forecasting the term structure of government bond yields. *Journal of Econometrics*, 130:337–364, 2006.
- [5] Greg Duffee. Forecasting with the term structure: The role of no-arbitrage restrictions. Economics Working Paper Archive 576, The Johns Hopkins University, Department of Economics, January 2011.
- [6] Gregory R. Duffee. Term premia and interest rate forecasts in affine models. *Journal of Finance*, 57(1):405–443, 02 2002.
- [7] Darrell Duffie and Rui Kan. A yield-factor model of interest rates. *Mathematical Finance*, 6(4):379–406, 1996.

- [8] Eugene F Fama and Robert R Bliss. The information in long-maturity forward rates. *American Economic Review*, 77(4):680–92, September 1987.
- [9] Scott Joslin, Kenneth J. Singleton, and Haoxiang Zhu. A new perspective on gaussian dynamic term structure models. *Review of Financial Studies*, 24(3):926–970, 2011.
- [10] Robert B. Litterman and Josè Scheinkman. Common Factors Affecting Bond Returns. *The Journal of Fixed Income*, 1(1):54–61, 1991.
- [11] Charles R Nelson and Andrew F Siegel. Parsimonious modeling of yield curves. *The Journal of Business*, 60(4):473–89, October 1987.
- [12] Giovanni Petris, Sonia Petrone, and Patrizia Campagnoli. *Dynamic Linear Models with R*. useR! Springer-Verlag, New York, 2009.
- [13] Oldrich Vasicek. An equilibrium characterization of the term structure. *Journal of Financial Economics*, 5(2):177–188, November 1977.